(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号

特開平１０－９７２８５

(43)公開日　平成１０年（１９９８）４月１４日

(51)Int.Cl.[6]	識別記号	庁内整理番号	ＦＩ		技術表示箇所
G10L 3/00	561		G10L 3/00	561 G	
	531			531 E	

審査請求　未請求　請求項の数９　ＯＬ　（全１４頁）

(21)出願番号　特願平８－２５１３７３

(22)出願日　平成８年（１９９６）９月２４日

(71)出願人　０００００６０１３
三菱電機株式会社
東京都千代田区丸の内二丁目２番３号

(72)発明者　岩▲崎▼　知弘
東京都千代田区丸の内二丁目２番３号　三菱電機株式会社内

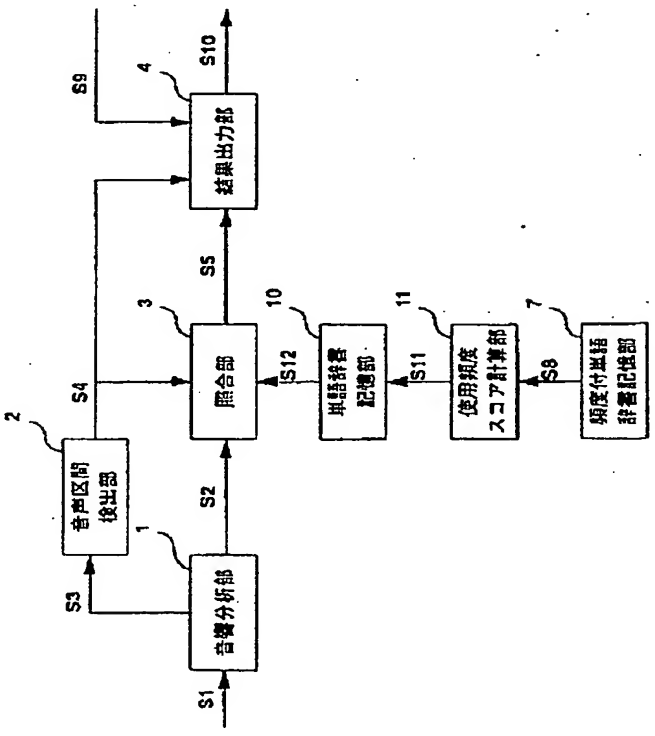
(74)代理人　弁理士　宮田　金雄　（外３名）

(54)【発明の名称】音声認識装置

(57)【要約】

【課題】　音声認識装置において、大語彙の音声が入力される場合でも高い精度で音声を認識する。

【解決手段】　頻度付単語辞書記憶部７の使用頻度を表す情報から使用頻度の高いものほど小さな値となる使用頻度スコアを計算し、単語辞書記憶部１０に記憶する使用頻度スコア計算部１１を備え、照合部３において入力された音声信号Ｓ１と単語辞書Ｓ１２が音響的にどの程度近いかを示す音響スコアに単語辞書記憶部に１０に記憶されているその単語の使用頻度スコアを規定の割合で加算して距離値Ｓ５とする。かくして、使用頻度が高い単語ほど小さな値となる使用頻度スコアを与え、パターン照合時にこれを加えることにより、使用頻度の高い単語の認識率できる。



1

【特許請求の範囲】

【請求項１】　入力される音声信号を一定時間毎に音響分析し、特徴パラメータベクトルと音声信号のパワーとに順次変換し出力する音響分析部と、当該音響分析部から受け取る上記音声信号のパワーの変化により上記音声信号の音声区間を検出し、当該音声区間の検出状況の変化により音声区間検出信号を出力する音声区間検出部と、上記音声区間検出信号の指示に従い上記音響分析部から出力される上記特徴パラメータベクトルの時系列のうち上記音声区間の部分と単語辞書記憶部に記憶されている単語辞書との照合を行い、入力された上記音声信号と上記単語辞書とのパターン照合を行い、距離値として出力する照合部と、上記音声区間検出信号により指示された時、既に受け取った上記距離値をソーティングして上記距離値の小さな１つ又は複数の単語を認識結果として出力する結果出力部とを有する音声認識装置において、認識対象単語の標準パターンと単語の使用頻度を表す情報を収める頻度付単語辞書記憶部と、当該頻度付単語辞書記憶部の使用頻度を表す情報から使用頻度の高いものほど小さな値となる使用頻度スコアを計算し、上記単語辞書記憶部に記憶する使用頻度スコア計算部とを備え、上記照合部において入力された上記音声信号と単語辞書が音響的にどの程度近いかを示す音響スコアに上記単語辞書記憶部に記憶されているその単語の上記使用頻度スコアを規定の割合で加算して距離値とすることを特徴とする音声認識装置。

【請求項２】　上記使用頻度スコア計算部は上記使用頻度スコアが規定の下限値より小さくならないように設定することを特徴とする請求項１に記載の音声認識装置。

【請求項３】　既存のデータベースから同じふり仮名を持つ単語の頻度を計数し、全体に対する上記単語の頻度を使用頻度とみなす使用頻度推定部を備えることを特徴とする請求項１に記載の音声認識装置。

【請求項４】　上記使用頻度推定部はふり仮名をローマ字表記したものが「ＯＵ」を含む時、規定の割合で「ＯＯ」に置き換えた読みがされるとし、元の上記単語の使用頻度を上記規定の割合で減ずると共に、新たに上記「ＯＵ」を上記「ＯＯ」で置き換えた単語を加え、当該新たな単語の使用頻度を元の上記単語の使用頻度の上記規定の割合とすることを特徴とする請求項３に記載の音声認識装置。

【請求項５】　上記使用頻度推定部はふり仮名をローマ字表記したものが「ＥＩ」を含む時、規定の割合で「ＥＥ」に置き換えた読みがされるとし、元の上記単語の使用頻度を上記規定の割合で減ずると共に、新たに上記「ＥＩ」を上記「ＥＥ」で置き換えた単語を加え、当該新たな単語の使用頻度を元の上記単語の使用頻度の上記規定の割合とすることを特徴とする請求項３に記載の音声認識装置。

【請求項６】　上記使用頻度推定部は任意の規定の割合で母音及び撥音が長音化されるものとし、元の上記単語の使用頻度を上記規定の割合で減ずると共に、新たに上記母音及び撥音を長音化したもので置き換えた単語を加え、当該新たな単語の使用頻度を元の上記単語の使用頻度の規定の割合とすることを特徴とする請求項３に記載の音声認識装置。

2

【請求項７】　上記使用頻度推定部は任意の規定の割合で音節毎に区切られるものとし、元の単語の使用頻度を上記規定の割合で減ずると共に、上記音節毎に区切られた単語を加え、当該新たな単語の使用頻度を元の上記単語の使用頻度の規定の割合とすることを特徴とする請求項３に記載の音声認識装置。

【請求項８】　上記使用頻度推定部は任意の規定の割合で促音が「つ」と発声されるものとし、元の単語の使用頻度を上記規定の割合で減ずると共に、促音が「つ」と発声された単語を加え、当該新たな単語の使用頻度を元の上記単語の使用頻度の上記規定の割合とすることを特徴とする請求項３に記載の音声認識装置。

【請求項９】　上記使用頻度推定部は規定の基準においてデータベースの内容を分類して、それぞれの分類毎に使用頻度を推定し、上記使用頻度スコア計算部は上記分類毎に使用頻度スコアを演算し、また同じ規定の基準において分類された話者の音声信号より学習された標準パターンをもって、未知の話者の音声信号の話者識別を行い上記話者がどの分類に近いかを示す話者認識別スコアを出力する話者識別部を備え、上記照合部は上記話者識別スコアと当該分類における単語の使用頻度スコアと単語の音響スコアを任意の規定の割合で加算して照合結果とすることを特徴とする請求項３に記載の音声認識装置。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は音声認識装置に関し、大語彙を単語を対象して認識するものに適用し得る。

【０００２】

【従来の技術】住所や姓名のような大語彙を対象とする音声認識装置では、類似する単語が多くなるため認識性能が低下し、また大語彙とのパターン照合を行うため演算量が膨大となる問題があり、実現が極めて難しいものである。従来、大語彙を対象とするこの種の音声認識装置として、特開平３－８４６００号公報に開示されたものを図９に示す。

【０００３】音響分析部１は入力される音声信号Ｓ１を一定時間毎に音響分析し、特徴パラメータベクトルＳ２と音声信号のパワーＳ３に変換し出力する。音声区間検出部２は音響分析部１から受け取る音声信号のパワーＳ３の変化により音声信号の音声区間を検出し、音声区間の検出状況の変化により音声区間検出信号Ｓ４を出力する。照合部３は音声区間検出信号Ｓ４の指示に従い、音

響分析部１から受け取る特徴パラメータベクトルＳ２の時系列のうち音声区間のものと単語辞書記憶部５から読み出される順番で単語辞書Ｓ５との照合を行い、入力された音声信号Ｓ１と単語辞書Ｓ６が音響的にどの程度近いかを示す音響スコアを距離値Ｓ５として順次出力する。

【０００４】なお頻度付単語辞書記憶部７は認識対象単語の読みを表すラベルと使用頻度を表す情報を収め、単語辞書ソーティング部６は頻度付単語辞書記憶部７の使用頻度の高い順番に単語情報を並び替え、単語辞書記憶部５は並び替えた単語情報を記憶する。また図中、Ｓ６は単語辞書、Ｓ７は単語辞書、Ｓ８は頻度付き単語辞書である。結果出力部４は音声区間検出信号Ｓ４又は外部から入力される出力要求信号Ｓ９が入力された時、既に受け取った距離値Ｓ５のうちまだ出力していないものを距離値Ｓ５によりソーティングして距離値Ｓ５の小さな１つ又は複数の単語を認識結果Ｓ１０として出力する。

【０００５】このような構成の音声認識装置の動作について説明する。認識に先だって単語辞書ソーティング部６では頻度付単語辞書記憶部７の内容を読み出し、使用頻度によりソーティングを行い、使用頻度の高い順番に単語辞書記憶部５に収める。以下、認識時の動作について説明する。認識装置は10m秒程度の時間を単位として処理が進められる。この単位時間をフレーム呼ぶ。音響分析部１はフレーム毎に入力された音声信号Ｓ１を音響分析し、特徴パラメータベクトルＳ２と音声信号のパワーＳ３に変換する動作を繰り返す。音響分析の手法としては、例えばＬＰＣ（Linear Prediction coefficient）分析やＦＦＴ（高速フーリエ変換）、フィルタバンクによる手法等が用いられる。

【０００６】次に音声区間検出部２の動作を説明する。音声区間検出部２では音声区間の検出は音声信号のパワーＳ３を監視し、音声信号のパワーＳ３がある閾値を越えたら音声区間の始端とし、閾値より下回ったら音声区間の終端候補とし、そのまま閾値以下で一定時間継続すると終端候補が正しかったものとして終端確定する。この時間は一般的には 0.3秒程度が適当とされている。0.3秒以内に再び閾値を越えて立ち上がると、先ほど検出した終端候補を無効とする。

【０００７】具体的に例をあげて音声区間検出部２の動作を、図１０を用いて説明する。図１０において「ほった」という発声の音声信号のパワーの変化の一例を示す。横軸は時間、縦軸は音声信号のパワーの大きさを表す。フレームＴ１からフレームＴ２までが「ほ」、フレームＴ２からフレームＴ３までが「っ」、フレームＴ３からフレームＴ４までが「た」の発声区間を想定している。図１０では音声信号のパワーは雑音レベルからフレームＴ１時点で閾値Ｐ１を越えフレームＴ２で下回る。再びフレームＴ３で閾値Ｐ１を越えフレームＴ４で下回る。フレームＴ５はフレームＴ４から 0.3秒経過した時点を指す。「っ」は音響的に促音に分類される。普通の発声では促音は 0.3秒以下の時間長となるため、この例でもフレームＴ２とフレームＴ３の間は 0.3秒以下の時間とする。上述の音声区間検出部２の動作に従えば、フレームＴ１からフレームＴ４を音声区間として検出する。

【０００８】音声区間検出部２では音声区間検出信号Ｓ４として始端信号、終端候補信号、終端確定信号の３種類を送出する。図１０ではフレームＴ１とフレームＴ３で始端信号を、フレームＴ２とフレームＴ４で終端候補信号を、フレームＴ５で終端確定信号を送出する。終端候補信号の後、終端確定信号が送出されずに、始端信号が送出された場合は、その前の終端候補信号の終端候補、つまりフレームＴ２を無効とすることを示す。

【０００９】照合部３では音響分析部１から音声区間の特徴パラメータベクトルＳ２が送られてくるので、音声区間検出信号Ｓ４で指定される始端信号から終端確定信号までの間、内部に蓄える。音声区間検出部２から音声区間検出信号Ｓ４として終端候補信号を受信したならばパターン照合を開始する。図１０にパターン照合を行うフレームを斜線で示す。パターン照合の方法はさまざまあるが、例えばＤＰ（Dynamic Programming）マッチングやＨＭＭ（Hidden Markov Model）による方法が適用できる。照合部３は単語辞書記憶部５の単語辞書Ｓ６を並べられている順番に読み出し、内部に蓄えられているフレームＴ１からフレームＴ２の区間の特徴パラメータベクトルＳ２に対するパターン照合を行い、距離値Ｓ５を結果出力部４に送出する。単語辞書記憶部５の中では単語辞書Ｓ６は頻度の高い順番に並べられているため、パターン照合は頻度の高い単語から行われることになる。

【００１０】続いてフレームＴ３で音声区間検出部２から音声区間検出信号Ｓ４として始端信号を受信した時には、フレームＴ１からフレームＴ２までの区間が無効であるとしパターン照合を中止する。さらに続いてフレームＴ４で音声区間検出部２から音声区間検出信号Ｓ４として終端候補信号を受信した時、内部に蓄えているフレームＴ１からフレームＴ４の区間の特徴パラメータベクトルＳ２に対するパターン照合を行い、入力された音声信号Ｓ１と単語辞書Ｓ６がどの程度近いかを示す音響スコアとその単語を距離値Ｓ５として順次、結果出力部４に出力する。

【００１１】結果出力部４では照合部３から送られてくる距離値Ｓ５に対しその音響スコアにより順次ソーティングを行う。音声区間検出部２からの音声区間検出信号Ｓ４として始端信号を受けた時、それまでにソーティングされた距離値Ｓ５をクリアする。音声区間検出部２からの音声区間検出信号Ｓ４として終端確定信号を受けた時、それまでにソーティングされた距離値Ｓ５のうち上位から１つ又は複数の認識結果Ｓ１０を出力する。この

5

出力結果を画面に出力したものを発声者が確認し、正しい認識結果が含まれていない場合には出力要求信号Ｓ９を入力する。この出力要求信号Ｓ９が入力された場合には、その時点までにソーティングされた距離値Ｓ５のうちで、まだ出力していないもののうち上位から１つ又は複数の単語を認識結果Ｓ１０として出力する。このなかに認識結果が含まれていない場合には、さらに上記のシーケンスが繰り返される。

【００１２】結果出力部４の処理の流れを図１０を用いてさらに説明する。フレームＴ１で結果出力部４は内部のデータをクリアする。フレームＴ２から照合部３から距離値Ｓ５が転送されてくるので順次その距離値Ｓ５によりソーティングを行い内部に蓄える。フレームＴ３でソーティングした結果をクリアする。フレームＴ４で再び照合部３から距離値Ｓ４が転送されてくるので順次その音響スコアによりソーティングを行い内部に蓄える。フレームＴ５でソーティングされた距離値Ｓ５のうち上位から１つ又は複数の単語を認識結果Ｓ１０として出力する。図１０に認識結果Ｓ１０を出力している期間を黒く塗った長方形で示している。フレームＴ５以降も照合部３から距離値Ｓ５が転送されてくるので順次ソーティングを行い内部に蓄える。フレームＴ６で外部から出力要求信号Ｓ９が入力されるので、その時点までソーティングされた距離値Ｓ５のうち上位から１つあるいは複数の単語を認識結果Ｓ１０として出力する。

【００１３】上記のように、従来の技術による音声認識装置では、大語彙の単語認識を行う場合に頻度の高い単語の認識結果は発声終了後 0.3秒で出力される。また、頻度の低い単語もしばらく後に装置に出力要求信号Ｓ９を送ることで認識結果を得ることができる。

【００１４】

【発明が解決しようとする課題】ところが従来の音声認識装置は以上のように構成されているので、どのように丁寧に発声しても頻度の低い単語は発声終了後 0.3秒以内では認識できない。また、姓名のように数万単語という大語彙を認識しようとする場合、「大野／小野」、「佐藤／里」など類似した単語が増加するため、認識率が低下するという問題があった。図１１に電話回線を通して収集した姓を発声する音声データに対する従来の音声認識装置の認識性能を示す。図中縦軸は誤り率、横軸はパターン照合に用いた頻度の高い単語数を対数（log ）で示す。図中実線で示したものが従来の音声認識装置による誤り率であり、点線はパターン照合に用いた単語辞書記憶部５の単語のなかに正解が含まれていなかった割合を示す。

【００１５】日本人の姓の総数は約58,000単語であり、右に行くほど多くの単語とパターン照合を行っている。頻度の高い 1,000単語を対象に認識する場合は、正解がこの1,000単語の中に含まれない割合である脱落率が30.7 唐ﾆ誤認識の16.5％を合わせて47.2％の発声が不正解

6

となるが、さらに時間をかけて照合を行い57,711単語を照合させた場合、 1.3％の脱落と62.1％の誤認識を合わせて、63.4％が不正解となることを示している。つまり、単語数を増やすと極端に認識率が低下し、結果として不正解が増加する問題があった。

【００１６】この発明は上記のような問題点を解消するためになされたもので、大語彙の場合でも高い精度で音声を認識し得る音声認識装置を提供するものである。

【００１７】

【課題を解決するための手段】この発明に係る音声認識装置は、入力される音声信号を一定時間毎に音響分析し、特徴パラメータベクトルと音声信号のパワーとに順次変換し出力する音響分析部と、その音響分析部から受け取る音声信号のパワーの変化により音声信号の音声区間を検出し、その音声区間の検出状況の変化により音声区間検出信号を出力する音声区間検出部と、音声区間検出信号の指示に従い音響分析部から出力される特徴パラメータベクトルの時系列のうち音声区間の部分と単語辞書記憶部に記憶されている単語辞書との照合を行い、入力された音声信号と単語辞書とのパターン照合を行い、距離値として出力する照合部と、音声区間検出信号により指示された時、既に受け取った距離値をソーティングして距離値の小さな１つ又は複数の単語を認識結果として出力する結果出力部とを有する音声認識装置において、認識対象単語の標準パターンと単語の使用頻度を表す情報を収める頻度付単語辞書記憶部と、その頻度付単語辞書記憶部の使用頻度を表す情報から使用頻度の高いものほど小さな値となる使用頻度スコアを計算し、単語辞書記憶部に記憶する使用頻度スコア計算部とを備え、照合部において入力された音声信号と単語辞書が音響的にどの程度近いかを示す音響スコアに単語辞書記憶部に記憶されているその単語の使用頻度スコアを規定の割合で加算して距離値とするものである。

【００１８】また次の発明に係る音声認識装置は、使用頻度スコア計算部は使用頻度スコアが規定の下限値より小さくならないように設定するものである。

【００１９】また次の発明に係る音声認識装置は、既存のデータベースから同じふり仮名を持つ単語の頻度を計数し、全体に対する単語の頻度を使用頻度とみなす使用頻度推定部を備えるものである。

【００２０】また次の発明に係る音声認識装置は、使用頻度推定部はふり仮名をローマ字表記したものが「ＯＵ」を含む時、規定の割合で「ＯＯ」に置き換えた読みがされるとし、元の単語の使用頻度を規定の割合で減ずると共に、新たに「ＯＵ」を「ＯＯ」で置き換えた単語を加え、その新たな単語の使用頻度を元の単語の使用頻度の規定の割合とするものである。

【００２１】また次の発明に係る音声認識装置は、使用頻度推定部はふり仮名をローマ字表記したものが「ＥＩ」を含む時、規定の割合で「ＥＥ」に置き換えた読み

がされるとし、元の単語の使用頻度を規定の割合で減ずると共に、新たに「ＥＩ」を「ＥＥ」で置き換えた単語を加え、その新たな単語の使用頻度を元の単語の使用頻度の規定の割合とするものである。

【００２２】また次の発明に係る音声認識装置は、使用頻度推定部は任意の規定の割合で母音及び撥音が長音化されるものとし、元の単語の使用頻度を規定の割合で減ずると共に、新たに母音及び撥音を長音化したもので置き換えた単語を加え、その新たな単語の使用頻度を元の単語の使用頻度の規定の割合とするものである。

【００２３】また次の発明に係る音声認識装置は、使用頻度推定部は任意の規定の割合で音節毎に区切られるものとし、元の単語の使用頻度を規定の割合で減ずると共に、音節毎に区切られた単語を加え、その新たな単語の使用頻度を元の単語の使用頻度の規定の割合とするものである。

【００２４】また次の発明に係る音声認識装置は、使用頻度推定部は任意の規定の割合で促音が「つ」と発声されるものとし、元の単語の使用頻度を規定の割合で減ずると共に、促音が「つ」と発声された単語を加え、その新たな単語の使用頻度を元の単語の使用頻度の規定の割合とするものである。

【００２５】また次の発明に係る音声認識装置は、使用頻度推定部は規定の基準においてデータベースの内容を分類して、それぞれの分類毎に使用頻度を推定し、使用頻度スコア計算部は分類毎に使用頻度スコアを演算し、また同じ規定の基準において分類された話者の音声信号より学習された標準パターンをもって、未知の話者の音声信号の話者識別を行い話者がどの分類に近いかを示す話者識別スコアを出力する話者識別部を備え、照合部は話者識別スコアとその分類における単語の使用頻度スコアと単語の音響スコアを任意の規定の割合で加算して照合結果とするものである。

【００２６】

【発明の実施の形態】以下図面を参照しながら、この発明の実施の形態を説明する。

【００２７】実施の形態１．図９との対応する部分に同一符号を付けた図１に、この発明による実施の形態１の

$$S(w) = -1.0 \times \log(P(w))$$

【００３２】式（１）においてｗは単語、Ｐ（ｗ）は単語ｗの使用頻度を確率で表したもの、Ｓ（ｗ）は単語ｗの使用頻度スコアである。Ｓ（ｗ）は使用頻度が高い単語には小さな値、使用頻度の大きな単語には大きな値となる。しかし、あまりに使用頻度が小さすぎると使用頻度スコアが非常に大きな値となり、どんなに丁寧に発声しても結果出力部４において上位の認識結果として出力されなくなるため、使用頻度スコアの下限値を設けることにより、非常に使用頻度の少ない単語でも、音響スコアが小さければ認識結果として出力することが可能となる構成としても良い。

音声認識装置を示す。図９について上述した従来の音声認識装置と同様に、音響分析部１は入力される音声信号Ｓ１を一定時間毎に音響分析し、特徴パラメータベクトルＳ２と音声信号のパワーＳ３に変換し出力する。音声区間検出部２は音響分析部１から受け取る音声信号のパワーＳ３の変化により音声信号Ｓ１の音声区間を検出し、音声区間の検出状況の変化により音声区間検出信号Ｓ４を出力する。

【００２８】照合部３は音声区間検出信号Ｓ４の指示に従い音響分析部１から受け取る特徴パラメータベクトルの時系列のうち音声区間のものと、単語辞書記憶部１０から読み出される順番でスコア付単語辞書Ｓ１２との照合を行い、入力された音声信号Ｓ１とスコア付単語辞書Ｓ１２がどの程度近いかを示す音響スコアと使用頻度スコアをある規定の割合で加え距離値Ｓ５として順次出力する。ここでこの実施の形態１の場合、頻度付単語辞書記憶部７は認識対象単語の読みを表すラベルと使用頻度を表す情報を収め、使用頻度スコア計算部１１は頻度付単語辞書記憶部７の使用頻度を表す情報に従い頻度付き単語辞書Ｓ８に使用頻度スコアを付加し、使用頻度の高い順番にスコア付単語辞書Ｓ１１として出力する。単語辞書記憶部１０は使用頻度の高い順番にスコア付単語辞書Ｓ１１を記憶する。

【００２９】結果出力部４は音声区間検出信号Ｓ４又は外部から入力される出力要求信号Ｓ９が入力された時、既に受け取った距離値Ｓ５のうちまだ出力していないものをソーティングして距離値Ｓ５の小さな１つ又は複数の単語を認識結果Ｓ１０として出力する。

【００３０】この様な構成の音声認識装置の動作について説明する。認識に先立って、使用頻度スコア計算部１１では頻度付単語辞書記憶部７の内容を読みだし、使用頻度からスコア付単語辞書Ｓ１１を求め、使用頻度の高い順番に単語辞書記憶部１０に記憶する。この使用頻度スコアの与えかたとしては、例えば次式のような演算式により求める方法がある。

【００３１】

【数１】

$$\cdots\cdots \quad (1)$$

【００３３】この音声認識装置の認識時の動作について説明する。音響分析部１、音声区間検出部２、結果出力部４の動作は、図９～図１１について説明した従来の音声認識装置と同様のためここでは説明を省略する。ここではこの実施の形態１の特徴である照合部３の動作について説明する。照合部３では従来の音声認識装置と同様に単語辞書記憶部１０のスコア付単語辞書Ｓ１２を順番に読みだしパターン照合を行うが、次式に示すように音響スコアＤ（ｗ）に対し使用頻度スコアＳ（ｗ）を重みＲで加える。

【００３４】

【数２】

$$D'(w) = D(w) + R \times S(w) \quad \cdots\cdots (2)$$

【００３５】これにより、使用頻度スコアの低い単語は認識しやすくし、スコアの高い単語は認識しづらくする。すなわち、使用頻度の高い単語は認識しやすくし、使用頻度の低い単語は認識しづらくする効果を与える。この実施の形態１によれば、図１１について上述した認識実験と同じ条件による認識実験を行った結果、57,711単語を認識対象語彙とした時の誤り率を63.4％から32.1％に改善できた。

【００３６】実施の形態２．上述の実施の形態１では、使用頻度スコア計算部１１の機能として使用頻度の高い順番に並べて単語辞書記憶部１０に収めるとして説明をしたが、H／Wが十分に速く全単語候補に対するパターン照合が高速に処理可能な場合や、加えて、「孤立単語音声認識における全探索法・ビームサーチ法・Ａ＊探索法の比較」（平成８年度春季日本音響学会講演論文集、２－５－１０、伊田正樹、中川聖一著）に記載されているビームサーチ法や枝刈り法に代表される演算量削減策を講じることにより、図１０におけるフレームＴ５までにパターン照合で全候補の照合結果を得られる場合には、従来の音声認識装置のように分割してパターン照合を行う必要はなく、加えて使用頻度スコア計算部１１の機能として使用頻度の高い順番に並べて単語辞書記憶部７に収める必要はない。

【００３７】このような高速にパターン照合の処理が可能な照合部３を有する音声認識装置を実施の形態２に示す。この音声認識装置の構成は実施の形態１と同様なので、ここでは説明を省略する。このような構成の音声認識装置の動作について説明する。認識に先立って、使用頻度スコア計算部１１では頻度付単語辞書記憶部７の内容を読み出し、使用頻度からスコア付単語辞書Ｓ１１を求め、単語辞書記憶部１０に記憶する。単語辞書記憶部１０には使用頻度の高い順番に並べることは必要なくランダムに並べて良い。この使用頻度スコアの与え方としては、上述した実施の形態１と同様である。

【００３８】音響分析部１、音声区間検出部２の動作は、図９～図１１について上述した従来の音声認識装置と同様であり、ここでは説明を省略する。図２はこの実施の形態２に基づく音声認識装置の動作を説明するタイミングチャートである。以下この図２を用いて、照合部３及び結果出力４の動作について説明する。フレームＴ５より前の処理は従来の音声認識装置と同様である。この実施の形態２による照合部３では十分に処理能力が高いため、フレームＴ５以前にパターン照合の処理を終了している。そのためフレームＴ５において、結果出力部４は音声区間検出部２の音声区間検出信号Ｓ４としての終端確定信号により、照合部３から転送された距離値Ｓ５をソーティングして複合スコアの小さい１つ又は複数の単語の認識結果Ｓ１０として出力する。さらに外部からの出力要求信号Ｓ９があった場合には、先に出力した認識結果Ｓ１０を除いて、さらに距離値Ｓ５の小さい１つ又は複数の単語を認識結果Ｓ１０として出力する。

【００３９】実施の形態３．上述した実施の形態１、実施の形態２では単語の終端候補が定まった後、一単語づつパターン照合を行う方式の音声認識装置について述べてきたが、フレーム同期型パターン照合を行う照合部３を用いても、同様の効果を実現できる。フレーム同期型パターン照合は全単語辞書に対するパターン照合を同時に進めて行く方法である、一単語づつパターン照合を行う方法に比べ、ワークメモリ量は大きく増加するという欠点はあるが、音声入力と平行してパターン照合を行えるため、パターン照合を効率的に行えるという特徴を持つ。フレーム同期パターンパターン照合は、例えば「フレーム同期化、ビームサーチ、ベクトル量子化の統合によるＤＰマッチングの高速化」（電子通信学会論文誌Ｄ、Vol.J71-D,No.9,pp1650-1659、迫江博昭、藤井浩美、吉田和永、亘理誠夫共著）等に記述されている方法を用いる。

【００４０】このような音声認識装置の構成は実施の形態１の構成と同じであり、ここでは説明を省略する。この実施の形態３としての音声認識装置の動作について説明する。音響分析部１及び音声区間検出部２の動作は実施の形態２と同じなので、ここでは説明を省略する。照合部３及び結果出力部４の動作について、図３を用いて説明する。まず照合部３の動作について、照合部３は音声区間検出部２からの音声区間検出信号Ｓ４の１つである始端信号によりパターン照合処理を開始し、音声区間検出信号Ｓ４の１つである終端確定信号により動作を終了する。

【００４１】距離値Ｓ５は毎フレームにおいて照合部３から出力される。結果出力部４は音声区間検出信号Ｓ４の１つである終端候補信号により、終端候補のフレームの距離値Ｓ５をソーティングし、音声区間検出信号Ｓ４の１つである終端確定信号により距離値１１の小さな１つ又は複数の単語を認識結果Ｓ１０として出力する。図３には、フレームＴ２、Ｔ４の２つの終端候補信号があるが、フレームＴ５において出力するのは、フレームＴ４において得られた認識結果Ｓ１０である。このようにフレーム同期型のパターン照合を行うことにより、従来の音声認識装置では演算を行っていなかったフレームＴ１からフレームＴ２及びフレームＴ３からフレームＴ４においても照合部３の処理を行うことができ、効率的な音声認識装置を実現できる。

【００４２】実施の形態４．上述の説明では使用頻度が既知の単語について述べたが、音声認識装置をある程度運用すれば、使用頻度が得られる場合があるが、運用初期の段階では使用頻度を得ることは難しい場合が多い。

11

しかし、例えば自治体にある住民台帳や企業の持つ顧客データベースや社員データベースには、住所、姓名、電話番号、性別、年齢等が記録されている。そのため例えば、住民に対する情報サービスシステムなどでは、住民台帳の内容から単語の使用頻度が推定できる。つまり人口の割合の多い住所、姓名、電話番号等の単語は使用頻度は高いと推定する。企業の持つ顧客データベースや社員データベースに対しても同様の推定が可能である。この実施の形態４では、姓をひとつの例として使用頻度を推定する音声認識装置について説明する。

【００４３】この実施の形態４による音声認識装置の構成を図１との対応部分に同一符号を付けて図４に示す。図４において、音響分析部１、音声区間検出部２、照合部３、結果出力部４、頻度付単語辞書記憶部７、使用頻度スコア計算部１１、単語辞書記憶部１０は実施の形態３と同様なので、ここでは説明を省略する。図におい

12

て、データベース１２は住民の姓が含まれているデータベースであり、この中には姓に対してかな文字でふり仮名がふられているものとする。また使用頻度推定部１３はデータベース１２から名の頻度情報と読みを生成するものである。さらにＳ１３は姓情報、Ｓ１４は頻度付き単語辞書である。

【００４４】頻度付単語辞書記憶部７の推定方法について説明する。まずデータベース１２を検索して、同じふり仮名を持つ姓をひとつの単語ｗとして、単語ｗに対する人口Ｎ（ｗ）を調べる。異なる漢字であってもふり仮名が同じであれば同一の単語ｗとして計数する。そしてそのような姓を持つ人が多ければ、その姓の使用頻度も高いと推定し、使用頻度Ｐ（ｗ）を次式で求める。

【００４５】

【数３】

$$P(w) = N(w)/(ALLN) \quad \cdots\cdots (3)$$

【００４６】式（３）において、ALLNはそのデータベース１２に含まれる全人口である。単語の読みはデータベース１２に含まれる読みを用いる。その後のスコア付単語辞書Ｓ１１の作成方法や音響分析部１、音声区間検出部２、照合部３、結果出力部４の動作は実施の形態３に等しいのでここでは説明を省略する。

【００４７】実施の形態５．上述した実施の形態４では頻度付単語辞書記憶部７の単語の読みをデータベース１２に含まれるふり仮名を用いる例を述べたが、一般的にデータベース１２に含まれるふり仮名は書く時のかな文字がふられており、音声認識装置に入力される発声とは一致しない場合がある。例えばデータベース１２における「佐藤」のかな文字表記は「さとう」であるが、約80％の人はこれを「さとー」と長音で発声する。残り約15％の人は文字通り「さとう」と発声する。「さ、と、う」と区切って発声する人も存在する。これらは全て「佐藤」と音声認識すべきなので、自動的にこれらの単語を加え、使用頻度を推定することにより、認識率を向上させることが可能となる。

【００４８】この実施の形態５の音声認識装置の構成は実施の形態４の図４に等しいので、ここでは説明を省略する。ただし、図４の使用頻度推定部１３の動作は実施の形態４に示したものとは異なり、単語の変形規則を用いて同じふり仮名に対する異なる読み方の単語を追加する機能を持つ。以下この実施の形態５における使用頻度推定部１３の動作を説明する。図５はこの発明における使用頻度推定部１３の動作を示す流れ図である。図において処理は「ＳＴＡＲＴ」から始まり「ＥＮＤ」で終る。まず図中ステップＳＴ１においてローマ字表記で「ＯＵ」が含まれる単語に対しては、ステップＳＴ２において「ＯＵ」を「ＯＯ」に変えた単語を追加する。使用頻度はもとの単語の値に対し 0.8の倍率を乗ずる。もとの「ＯＵ」を含む単語の使用頻度は 0.2の倍率を乗じる。

【００４９】次にステップＳＴ３において、ローマ字表記で「ＥＩ」が含まれる単語に対しては、ステップＳＴ４において「ＥＩ」を「ＥＥ」に変えた単語を追加する。使用頻度は元の単語の確率に対し 0.7の倍率を乗ずる。元の「ＥＩ」を含む単語の使用頻度は 0.3の倍率を乗じる。またステップＳＴ５において、促音を含む単語であったなら人により促音を「つ」と発声することがあるため、ステップＳＴ６において促音を「つ」に変えた単語を追加する。使用頻度は元の単語の値に対し、0.05の倍率を乗じる。元の促音を含む単語の使用頻度は、0.95の倍率を乗じる。

【００５０】次にステップＳＴ７において、全単語に対し長音化した単語と切断化した単語を追加する。元の単語の使用頻度に対し長音化した単語は 0.1の倍率を乗じ、切断化した単語には0.05の倍率を乗じ使用頻度とする。元の単語の使用頻度は0.85の倍率を乗じて変更する。ただし、母音及び撥音の長音化においては最後の音節は長音化しない場合もあるため、このような変形規則を用いても良い。

【００５１】このような構成の音声認識装置による具体的な処理結果を示す。図６はあるデータベース１２を用いたときの実施の形態４に示された使用頻度推定部１３で推定される単語と使用頻度であるとする。これに対し、この実施の形態５における使用頻度推定部１３では図７に示す２０個の単語が推定される。図７において、ハイフン（－）は母音及び撥音が長音化されていることを示し、点（・）は音節が切断されていることを示す。母音及び撥音の長音化では、最期の音節は長音化しないという変形規則を用いている。

【００５２】各単語の使用頻度は、図５の流れに従い規定の倍率を掛けられている。例えば「あべ」は図５のステップＳＴ７の規則を適用されて母音が長音化された

13

「あーべ」と音節毎に切断された「あ・べ」が追加される。使用頻度は元の使用頻度 0.04598に対し「あべ」が0.85倍、「あーべ」が0.10倍、「あ・べ」が0.05倍されている。ただし、「にった」に関しては音節毎に切断された単語と、もとの単語が同じとなるため、「にった」の使用頻度が0.90倍されている。

【００５３】このようにこの実施の形態５によれば、データベース１２のふり仮名から、様々な発声の変形とその使用頻度を推定するため、良好な認識性能を示す音声認識装置を実現できる。なお上述の倍率の値は任意の調査結果から経験的に求めたものであるが、これらはデータベースに応じて変更しても良い。

【００５４】実施の形態６．データベースの内容において人口に偏りがある場合がある。例えば名前では男性、女性で人口が異なる。そのため、音声信号が男性か女性かの情報を追加すればさらに認識性能を高めることができる。この実施の形態６の音声認識装置の構成を図８に示す。図において、音響分析部１、音声区間検出部２、照合部３、結果出力部４、頻度付単語辞書記憶部７、使用頻度スコア計算部１１、単語辞書記憶部１０、データベース１２、使用頻度推定部１３の構成は、上述した実施の形態５と同様である。

【００５５】図８において話者識別部１４は音声区間検出部２からの音声区間検出信号Ｓ４に従い、音響分析部１からの特徴パラメータベクトルＳ２を比較し話者識別を行うものである。ここでは話者識別の対象を、男性、女性として、性別により姓名のうち名前を男性、女性で分類して記憶しておくことを一例として説明する。

14

【００５６】まず、認識に先立ち使用頻度推定部１３では、データベース１２の同じ名前に対し男性、女性に分けて別の単語として頻度付き単語辞書Ｓ１４を作成する。以下、使用頻度スコア計算部１１でも、別々の単語としてスコアを計算し、単語辞書記憶部１０に格納する。これにより、頻度付単語辞書記憶部７、単語辞書記憶部１０の記憶量は２倍になる。また話者識別部１４には話者識別のための標準パターンが入れられる。話者識別の方法としては多くのものが提案されているが、ここではベクトル量子化を用いる方法を一例として説明する。

【００５７】話者識別部１４には男性用を１、女性用を２としてそれぞれM個の標準パターンを用意する。この標準パターンは男性、女性それぞれの音声信号からＬＢＧ（Linde Buzo Gray ）アルゴリズム等を用いて学習される。性別ｉのm番目の標準パターンをＭ（ｉ，ｍ）、フレームｔの特徴パラメータベクトル９をＬ（ｔ）とすると、話者識別スコア２７であるＳ２（ｉ）は、次式の演算式で求められる。

【００５８】

【数４】

$$S2(i) = \frac{\sum_{t=T1}^{T4} min(dis(M(i,m), L(t)))_{m=1,M}}{T4 - T1 + 1} \quad \cdots\cdots (4)$$

【００５９】式（４）において、次式

【００６０】

【数５】

$$min(X(m))_{m=1,M} \quad \cdots\cdots (5)$$

【００６１】は要素Ｘ（ｍ）のｍ＝１、Ｍに関する最小値を意味する。また、

【００６２】

【数６】

$$dis(M(i,m), L(t)) \quad \cdots\cdots (6)$$

【００６３】はＭ（ｉ，ｍ）とＬ（Ｔ）の距離値を意味する。式（４）の演算はフレームＴ４においてまとめて行うことも可能であるし、フレームＴ１からフレーム同期的に行うことも可能である。フレームＴ１、Ｔ４は音声区間検出信号Ｓ４として通知される。このようにして得られた話者識別スコアＳ１５は照合部３においてＲ２の割合で音響スコアと使用頻度スコアに加えられ、男性用のものと女性用のものと小さなほうが最終的な照合結果となる。

【００６４】

【数７】

$$D'(w) = min(D(w) + R \times S1(i,w) + R2 \times S2(i))_{i=1,2} \quad \cdots\cdots (7)$$

【００６５】式（７）においてＤ（ｗ）、Ｒは式（２）で用いたものと同じであり、Ｓ１（ｉ，ｗ）は性別ｉの単語ｗに対する使用頻度スコアである。

【００６６】上述の説明では話者識別部１４の標準パターンを単語辞書記憶部１０の標準パターンとは別のものとして説明したが、男性用、女性用の単語辞書記憶部１０の標準パターンを持つマルチテンプレートの音声認識装置では、これを話者識別用に流用することも可能であ

り、このような構成でも上述と同様の効果を実現できる。また上述の説明では性別により話者識別を行う例を述べたが、年齢や日本人名か英語名かの言語などによりデータベース１２を分類して話者識別を行うことも可能であり、同様に効果を実現できる。

【００６７】

【発明の効果】以上のようにこの発明によれば、単語の使用頻度から計算した使用頻度スコアを音響スコアに規定の割合で加えて距離値を得るため、頻度の多い単語の認識性能を高めることができ、大語彙の場合でも全体として認識率を格段的に向上し得る音声認識装置を実現できる。

【００６８】また次の発明によれば、非常に使用頻度の低い単語のスコアの下限値を設けたため、極めて使用頻度の低い単語でも音響スコアが良好であれば、上位の認識結果とでき、かくするにつき、大語彙の場合でも全体として認識率を格段的に向上し得る音声認識装置を実現できる。

【００６９】また次の発明によれば、既存のデータベースから使用頻度を推定できるため、使用頻度が明確でない単語に対しても、使用頻度スコアを設定でき、かくするにつき、大語彙の場合でも全体として認識率を格段的に向上し得る音声認識装置を実現できる。

【００７０】また次の発明によれば、既存のデータベースのふり仮名にローマ字表記で「ＯＵ」を含む単語がある時、これを「ＯＯ」と変更した単語を追加し、使用頻度を規定の割合で設定するため、ふり仮名とは異なる発声をされた場合も認識でき、かくするにつき、大語彙の場合でも全体として認識率を格段的に向上し得る音声認識装置を実現できる。

【００７１】また次の発明によれば、既存のデータベースのふり仮名にローマ字表記で「ＥＩ」を含む単語がある時、これを「ＥＥ」と変更した単語を追加し、使用頻度を規定の割合で設定するため、ふり仮名とは異なる発声をされた場合も認識でき、かくするにつき、大語彙の場合でも全体として認識率を格段的に向上し得る音声認識装置を実現できる。

【００７２】また次の発明によれば、既存のデータベースのふり仮名に対し母音を長音化した単語を追加し、使用頻度を規定の割合で設定するため、ふり仮名とは異なる発声をされた場合も認識でき、かくするにつき、大語彙の場合でも全体として認識率を格段的に向上し得る音声認識装置を実現できる。

【００７３】また次の発明によれば、既存のデータベースのふり仮名に対し音節毎に区切られた単語を追加し、使用頻度を規定の割合で設定するため、ふり仮名とは異なる発声をされた場合も認識でき、かくするにつき、大語彙の場合でも全体として認識率を格段的に向上し得る音声認識装置を実現できる。

【００７４】また次の発明によれば、既存のデータベースのふり仮名に促音を含む単語がある時、これを「つ」と変更した単語を追加し、使用頻度を規定の割合で設定するため、ふり仮名とは異なる発声をされた場合も認識でき、かくするにつき、大語彙の場合でも全体として認識率を格段的に向上し得る音声認識装置を実現できる。

【００７５】また次の発明によれば、規定の基準においてデータベースの内容を分類して使用頻度を推定し、認識時には話者照合を行い、話者照合スコアを使用頻度スコアと音響スコアにある割合で加えるため、良好な認識性能を得ることができ、かくするにつき、大語彙の場合でも全体として認識率を格段的に向上し得る音声認識装置を実現できる。

【図面の簡単な説明】

【図１】　この発明による音声認識装置の実施の形態１の構成を示すブロック図である。

【図２】　この発明による音声認識装置の実施の形態２の動作の説明に供するタイミングチャートである。

【図３】　この発明による音声認識装置の実施の形態３の動作の説明に供するタイミングチャートである。

【図４】　この発明による音声認識装置の実施の形態４の構成を示すブロック図である。

【図５】　この発明による音声認識装置の実施の形態５の使用頻度推定部の動作を示すフローチャートである。

【図６】　この発明による音声認識装置の実施の形態４の使用頻度推定部の処理結果の説明に供する図表である。

【図７】　この発明による音声認識装置の実施の形態５の使用頻度推定部の処理結果の説明に供する図表である。

【図８】　この発明による音声認識装置の実施の形態６の構成を示すブロック図である。

【図９】　従来の音声認識装置の構成を示すブロック図である。

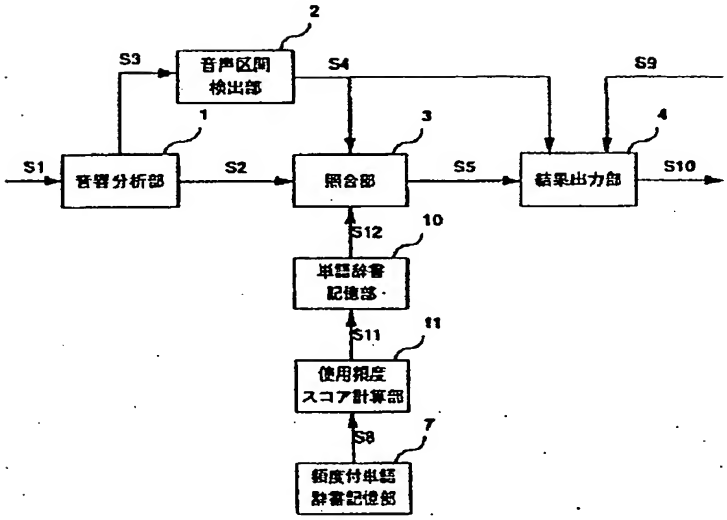
【図１０】　図９の音声認識装置における音声区間検出部の動作の説明に供するタイミングチャートである。

【図１１】　従来の音声認識装置による認識性能の説明に供する特性曲線図である。

【符号の説明】

１　音響分析部
２　音声区間検出部
３　照合部
４　結果出力部
５　単語辞書記憶部
６　単語辞書ソーティング部
７　頻度付単語辞書記憶部
１０　単語辞書記憶部
１１　使用頻度スコア計算部
１２　データベース
１３　使用頻度推定部
１４　話者識別部

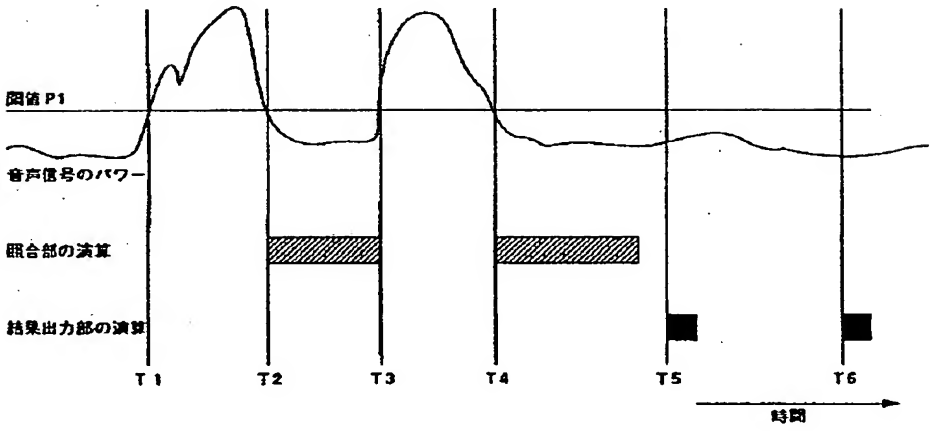
【図１】



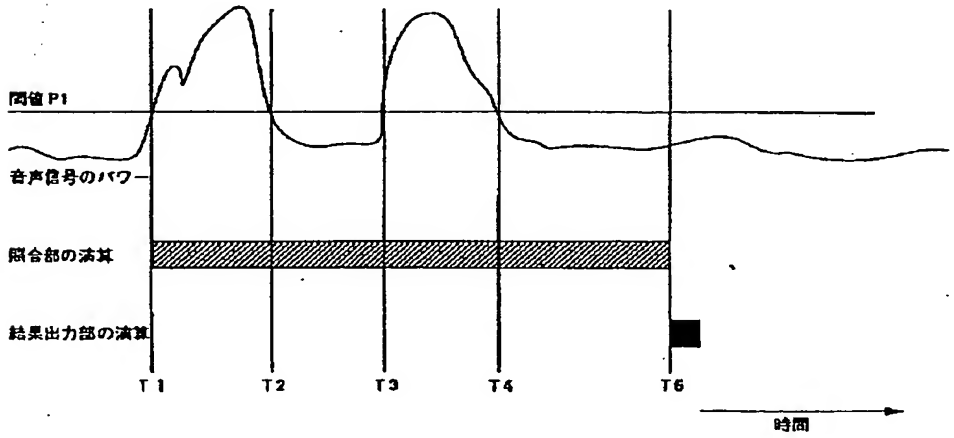
【図６】

単語	使用頻度
あべ	0.04598
さとう	0.04054
にった	0.00086
せいの	0.00076

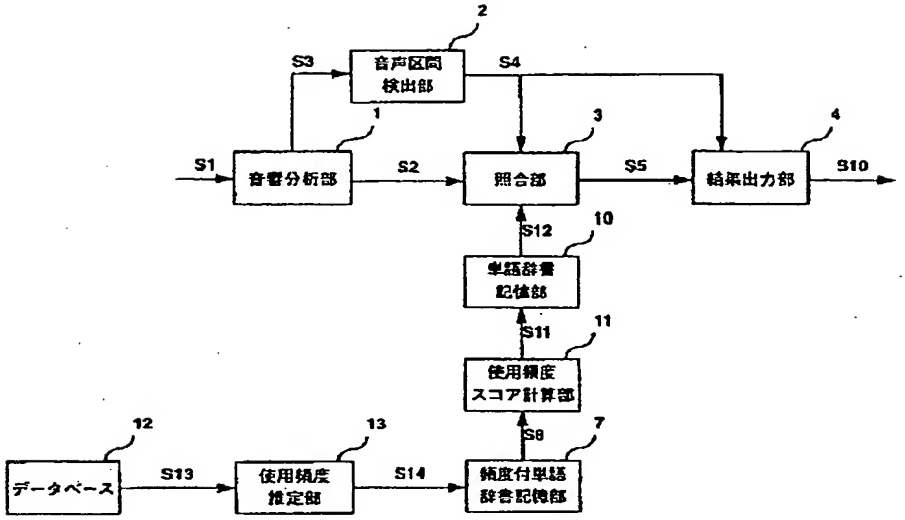
【図２】



【図３】



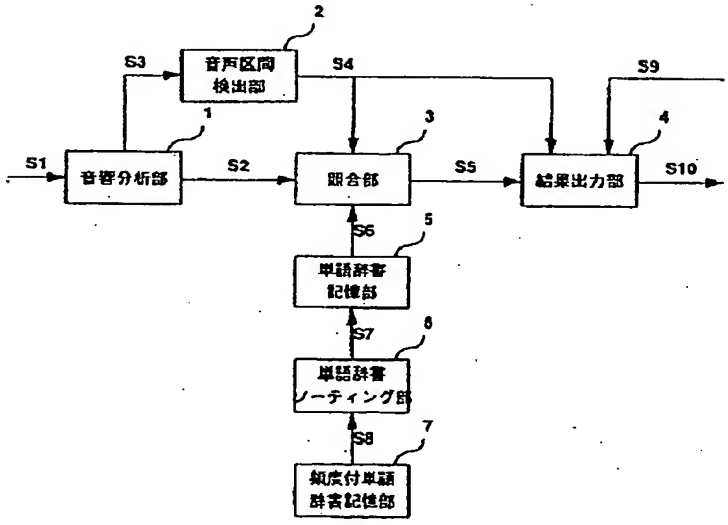
【図４】



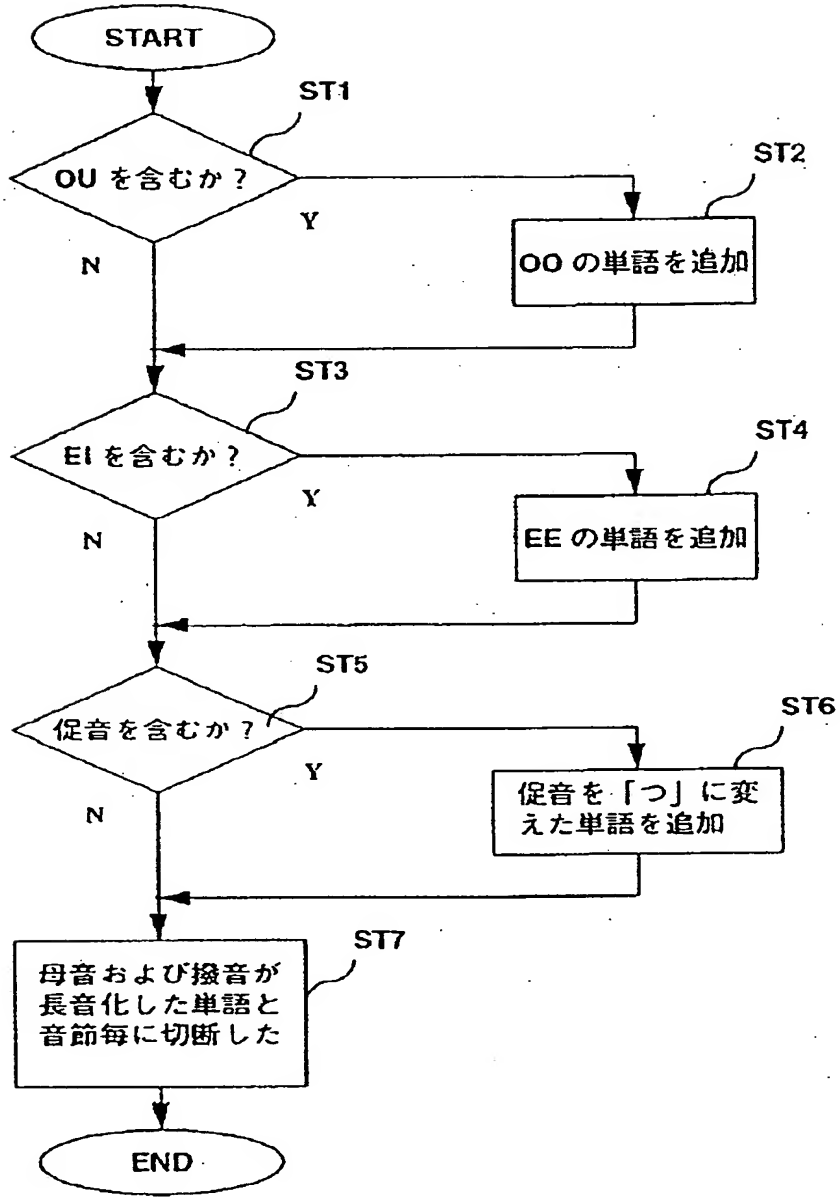
【図７】

単語	使用頻度
あべ	0.04598 × 0.85
あーべ	0.04598 × 0.10
あ・べ	0.04598 × 0.05
さとう	0.04054 × 0.2 × 0.085
さとお	0.04054 × 0.8 × 0.085
さーとーう	0.04054 × 0.2 × 0.010
さーとーお	0.04054 × 0.8 × 0.010
さ・と・う	0.04054 × 0.2 × 0.005
さ・と・う	0.04054 × 0.8 × 0.005
にった	0.00086 × 0.95 × (0.85 + 0.05)
につた	0.00086 × 0.05 × 0.85
にーった	0.00086 × 0.95 × 0.10
にーつーた	0.00086 × 0.05 × 0.10
に・つ・た	0.00086 × 0.05 × 0.05
せいの	0.00076 × 0.6 × 0.85
せえの	0.00076 × 0.4 × 0.85
せーいーの	0.00076 × 0.6 × 0.10
せーえーの	0.00076 × 0.4 × 0.10
せ・い・の	0.00076 × 0.6 × 0.05
せ・え・の	0.00076 × 0.4 × 0.05

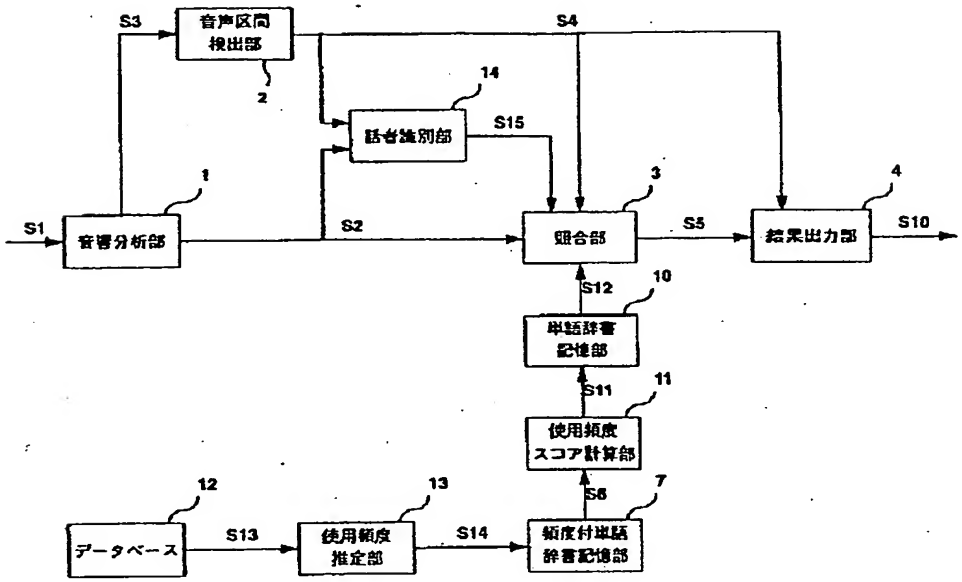
【図９】



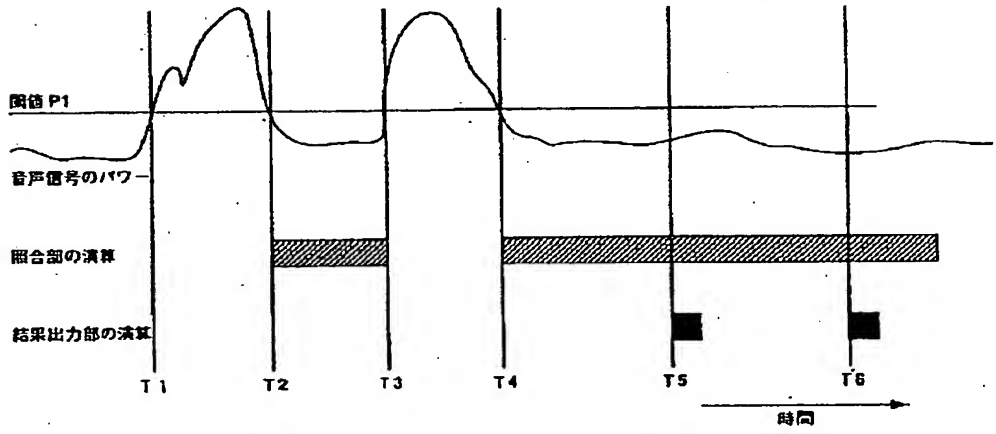
【図５】


START
ST1
OU を含むか？
Y
N
ST2
OO の単語を追加
ST3
EI を含むか？
Y
N
ST4
EE の単語を追加
ST5
促音を含むか？
Y
N
ST6
促音を「つ」に変えた単語を追加
ST7
母音および撥音が長音化した単語と音節毎に切断した
END

【図８】


S3
音声区間検出部
2
S4
14
話者識別部
S15
1
3
4
S1
音響分析部
S2
照合部
S5
結果出力部
S10
S12
10
単語辞書記憶部
S11
11
使用頻度スコア計算部
S6
7
12
13
データベース
S13
使用頻度推定部
S14
頻度付単語辞書記憶部

【図１０】


閾値 P1
音声信号のパワー
照合部の演算
結果出力部の演算
T1
T2
T3
T4
T5
T6
時間

【図１１】